

ENRICHMENT AND POPULATION OF A GEOSPATIAL ONTOLOGY FOR SEMANTIC INFORMATION EXTRACTION

M. Kokla^{1*}, V. Papadias², E. Tomai¹

¹ School of Rural and Surveying Engineering, National Technical University of Athens, Greece – (mkokla, etomai)@mail.ntua.gr

² Dept. of Geography, Harokopio University of Athens, Greece - veeqpap@gmail.com

Commission IV, WG IV/2

KEY WORDS: Ontology Development, Ontology Enrichment, Ontology Population, Semantic Information Extraction, Geospatial Ontology

ABSTRACT:

The massive amount of user-generated content available today presents a new challenge for the geospatial domain and a great opportunity to delve into linguistic, semantic, and cognitive aspects of geographic information. Ontology-based information extraction is a new, prominent field in which a domain ontology guides the extraction process and the identification of pre-defined concepts, properties, and instances from natural language texts. The paper describes an approach for enriching and populating a geospatial ontology using both a top-down and a bottom-up approach in order to enable semantic information extraction. The top-down approach is applied in order to incorporate knowledge from existing ontologies. The bottom-up approach is applied in order to enrich and populate the geospatial ontology with semantic information (concepts, relations, and instances) extracted from domain-specific web content.

1. INTRODUCTION

The massive amount of user-generated content available today presents a new challenge for the geospatial domain and a great opportunity to delve into linguistic, semantic, and cognitive aspects of geographic information. Although unstructured or semi-structured texts pose significant challenges due to the use of natural language, they provide a rich source of knowledge about places, events, phenomena, geospatial concepts, etc.

Place names have attracted a lot of research, since they play a central role in geographic information retrieval and call for elaborate approaches for their identification and disambiguation (Jones and Purves, 2008). Besides places, concepts are also important and a crucial element for describing the meaning of texts. The extraction of concepts from natural language texts may provide the basis for:

- revealing immanent spatial knowledge from text that can be formally described and further processed for semantic annotation of textual resources
- exploring how geographic concepts are described in natural language
- enabling semantic queries over textual resources overcoming the limitations of keyword-based search

The enrichment of unstructured natural language texts with machine readable semantic knowledge is necessary for supporting the semantic annotation, search, and retrieval of relevant web resources. Ontology-based information extraction (Karkaletsis et al., 2011) is a prominent field of information extraction in which an ontology is used as a means to formally describe domain knowledge and assist the extraction of pre-defined concepts, properties, and instances.

In own prior work (Kokla et al., 2018), a tool for the semantic information extraction and enrichment of natural language texts with spatial concepts and entities was described. The process was guided by a lightweight generic geospatial ontology developed based on Princeton WordNet (PWN) (Fellbaum, 1998). In general, the quality of the information extraction process is highly dependent on the sophistication of the underlying ontology. The more refined the ontology is, the more the wealth of knowledge that can be extracted grows.

The present paper describes the enrichment and population of the developed geospatial ontology using both a top-down and a bottom-up approach (Figure 1). A top-down approach is applied in order to extend the ontology by integrating knowledge from existing ontologies. A bottom-up approach is applied in order to enrich and populate the geospatial ontology with concepts and instances extracted from domain-specific web content.

In contrast to ontologies that capture authoritative or expert knowledge, an ontology for semantic information extraction from natural language texts needs to possess different characteristics. The ontology should include a wealth of geospatial concepts, abstract and concrete, concepts referring to places or natural and manmade spatial features, but also to geospatial primitives, spatial relations, natural and social processes, representation tools, etc.

The ontology should also have a natural language anchorage (Nedellec and Nazarenko, 2005) and include alternative linguistic realizations of the same entities (i.e., synonyms of ontology concepts and co-references of instances). For example, in order to be able to extract from natural language texts terms

* Corresponding author

such as temblor and seism and map them to their corresponding ontology concept earthquake, the ontology should include information on synonymous terms; otherwise these terms are missed during the information extraction process. The same holds for instances: different entity mentions (e.g., USA, US, United States) are linked to the the full-name form of the instance (United States of America) in order to be able to extract these entity mentions and map them to the single entity they refer to.

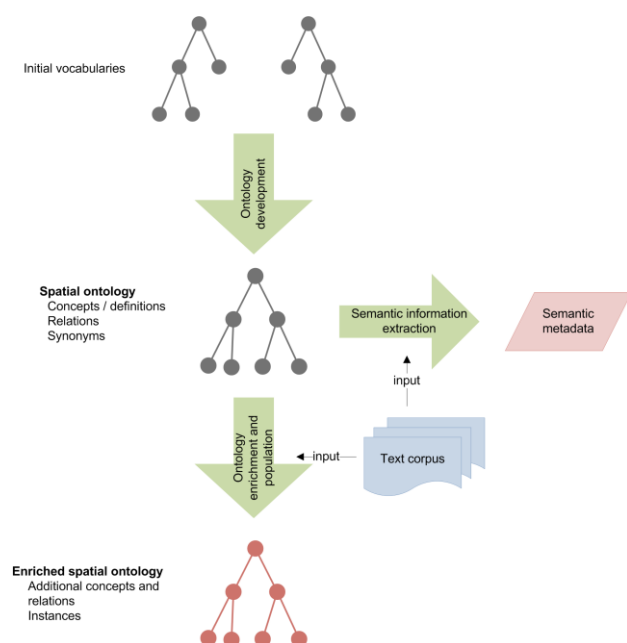


Figure 1. Workflow of the proposed approach

Furthermore, in contrast to authoritative ontologies which usually represent expert knowledge of the domain, ontologies for information extraction from natural language texts should represent a more commonsensical view of the domain. Usually this also involves the integration of different views of domain concepts and the inclusion of multiple relations among them.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature regarding ontology development, enrichment, and population. Section 3 outlines the approach for the top-down and bottom-up enrichment and population of a geospatial ontology. Finally, Section 4 draws conclusions and discusses future directions.

2. RELATED WORK

Ontology-based information extraction (OBIE) employs ontologies for relating conceptual knowledge (concepts) to its linguistic realizations (the terms that express that concepts) (Nedellec and Nazarenko, 2005). OBIE systems include several components which generally consist in: (1) a preprocessing module to convert the texts into a format that can be further processed, (2) an information extraction module that may employ different techniques, which are however always guided by an ontology, (3) an ontology generation module (Wimalasuriya and Dou, 2010). The output of such systems is the information extracted from text, which may be further used in other tasks such as semantic annotation and search.

Ontologies may either function as an input to guide the information extraction process or be generated by the information extraction process (Wimalasuriya and Dou, 2010). In the first case the ontology may be manually defined by domain experts, or imported from other sources. In the latter case the ontology may be automatically or semi-automatically generated from text, either from scratch or using an existing ontology as the basis. Ontology enrichment is defined as “the task of extending an existing ontology with additional concepts and semantic relations and placing them at the correct position in the ontology” whereas ontology population is defined as “the task of adding new instances of concepts to the ontology” (Petasis et al., 2011).

Ontology enrichment and population have attracted significant research efforts due to their importance in information extraction, semantic annotation and search. Paci et al. (2010) provide a method for automatically associating concepts of an ontology to their mentions in texts, after connecting the text mentions to Wikipedia articles, thus the semantic annotation performed is two-folded. A pipelined supervised learning approach for interlinking concepts mentions in documents to ontology concepts is presented in Melli and Ester (2010). The authors are able to identify mentions of concepts not yet present in the ontology and then implement tests that select candidate ontology concepts to be linked to these concept mentions. OntoPop (Amardeilh, 2016) is a methodology that maps linguistic extractions from text documents with ontology concepts based on knowledge acquisition rules. Moreover, the methodology has been used to implement the OntoPop platform for document annotation and ontology population, which incorporates an Information Extraction tool and a Knowledge Representation one respecting however their independency. Benammar et al. (2015) present an approach for populating the COSCHKR (Colour and Space in Cultural Heritage) ontology from Cultural Heritage scientific papers. After the pre-processing phase (co-reference resolution, sentence splitting, tokenization, and POS tagging), the authors build a gazetteer based on WordNet (Fellbaum, 1998), where concepts in COSCHKR are matched to their synonyms. After that, they perform triples extraction in terms of relations and concepts. Finally, they associate the triples with the ontology and then perform property matching with SPARQL queries.

In the geospatial domain, ontologies have been developed by different organizations for the formalization of geographic concepts. Ontologies have been used for the identification and disambiguation of place names in the context of geographic information retrieval (Jones et al., 2001; Lutz and Klien, 2006), for the spatialization (Cooper et al., 2015; Bruggmann and Fabrikant, 2014a, 2014b) and exploration of text corpora (Derungs and Purves, 2014), for semantic search (Hu et al., 2015), and for the extraction of place emotions (Ballatore and Adams, 2015) and spatiotemporal and semantic information on natural hazards (Wang and Stewart, 2015). However, to our knowledge there has been no attempt to enrich an existing ontology with concepts and instances using as input natural language texts.

3. ONTOLOGY ENRICHMENT AND POPULATION

The first task of ontology development deals with the specification of spatial concepts and their between relations. Focusing the analysis on initiatives in the context of education, Kavouras et al. (2016) have selected geospatial concepts based on a thorough analysis of different vocabularies and ontologies.

The METHONDOLOGY (Fernandez et al., 1997) process has been adopted for the development of the ontology which includes the processes of specification, knowledge acquisition, conceptualization, integration, and implementation. The resulting ontology includes concepts varying in complexity that are organised in a three-level hierarchy. Nineteen clusters of basic and generic notions and subject areas such as geometric primitives, spatial relations, space-time primitives, geography etc. constitute the first and second tiers of the ontology. In the third tier, more detailed and specific concepts are included enriched with definitions from Princeton WordNet (PWN) as well as synonyms from synsets that provide alternative realizations of these concepts (Fellbaum, 1998).

PWN lemmas have the form of synsets; synonym terms that constitute a lemma and differentiate it from other lemmas. Synsets of the concepts in the ontology provide alternative concept mentions. A total number of 290 terms comprise the set of synonyms, 49 of which are not unique; they appear in more than one relative synsets. Characteristic examples are: (a) state which appears in the synsets: “**country**, land, state”, “body politic, commonwealth, **country**, land, nation, res publica, state”, and “province, state” and (b) “**biotic community**, community”, “**community**”, and “community, **residential area**, residential district” (bold fonts denote the concept term in the ontology). This fact impedes the mapping of a term from the text to the proper ontology concept.

An important aspect of ontology development is the definition of interrelations among concepts. To discover these interrelations, the concepts have been further analysed based on PWN’s large lexical database. These have been first interlinked through semantic relations such as hyponymy/ hypernymy, and meronymy as extracted from PWN’s hierarchy.

3.1 Top-Down Ontology Enrichment

The process of ontology enrichment has been undertaken taking into account other resources such as BabelNet (Navigli and Ponzetto, 2012), Wikipedia and GeoNames (2018) to extend the number of interrelations, to validate the ontology structure and schema, and to differentiate between concepts and instances when necessary. In what follows, we describe the tasks undertaken to enrich and validate the ontology in terms of concepts and schema.

3.1.1 Hyponymy: Regarding hypernymic/ hyponymic relations, these comprise the majority of interrelations among concepts in the ontology. Since there is considerable critique towards PWN’s noun hierarchy (Gandemi et al. 2003, Guarino and Welty, 2004, Kaplan and Schubert, 2001, McCraey and Prangnawarat, 2016), to validate our ontology schema, further analysis of the definitions of the concepts as provided by PWN, has been pursued. This endeavor has indicated that in some cases there is a difference between the hypernym of a concept as provided by PWN’s hierarchy and the hypernym as given by the definition. For instance, earthquake’s hypernym from PWN schema is geological phenomenon, while the definition states: “shaking and vibration at the surface of the earth resulting from underground movement along a fault plane or from volcanic activity”.

This detailed analysis of the ontology concepts, showed that among the 221 definitions examined, 84 of them (38%) do not mention the same hypernym as the one indicated in the PWN hierarchy. This result had been taken into account leading us to

the following decisions: (a) changing the schema of the ontology, meaning that the hypernym derived from the definition was preferred from the one provided by the PWN hierarchy, (b) keeping the schema as such, by accepting the PWN hierarchy, while adding the hypernym extracted from the definition as alternative hypernym, (c) the same as b, but discarding completely the hypernym from the definition as marginal, or extreme, and (d) by providing two hypernyms for the concept in cases when the concept extracted from the definition was already part of the ontology.

Moreover, there are 67 concepts that have no direct hypernyms or hyponyms; they constitute “island concepts” in our ontology schema. Hypernymic relations are not defined for them because direct hypernyms are either very generic (PWN Tops) or not characteristic of the geospatial domain, and have been also left out. Moving downwards the hierarchy, most of these concepts do not have any hyponyms so such relation cannot be assigned to them. Examples of such concepts include cartography, coastline, global change, etc. These concepts need to be included in a geospatial ontology, even though they somehow violate the ontology schema by having no connections to other concepts.

3.1.2 Hyponymy: To partly solve the above-mentioned issue, we make once again use of other resources like BabelNet. In this way, we enrich the ontology with new relations, for instance for cartography, BabelNet includes four hyponyms (collaborative mapping, critical cartography, planetary cartography, terrain cartography). These can be used in the information extraction process in the form of gazetteers or can extend the ontology as new concepts. Another way to enrich our ontology, especially with hyponymic relations, is to refer to the GeoNames geographical database (2018), which includes nine feature classes with 645 feature codes as their subcategories. Although the GeoNames schema is quite flat, the plethora of feature codes can provide a basis for ontology enrichment especially in terms of natural and man-made features on the Earth’s surface. To establish these relations, an analysis of the definitions in GeoNames has been pursued with a parallel search in PWN to ground the process. For instance, the ontology concept body of water has 7 sub-concepts in the original version of the ontology: channel, lake, ocean, sea, stream, waterfall, and waterway. These were further enriched by GeoNames features: inlet (also part of lake and pond), lagoon (hyponym of lake), light, and pool.

3.1.3 Meronymic Relations: PWN does not document fully and consistently meronymic relations. However, our concepts are interlinked through these, in a limited number of cases. More specifically, 18 meronymic relations hold between 29 concepts. These include two meronym kinds (member and part). For instance, province is member of country, while hydrosphere is part of Earth. Meronymic relations, however, need further examination. Again, alternative sources of information were considered to provide this type of relation. For instance, line, the lemma in PWN defined as “a length (straight or curved) without breadth or thickness; the trace of a moving point”. Whereas, in BabelNet, line has part the concept point, and is also part of the concept plane, both concepts (point and plane) are included in the ontology, so meronymic relations can be introduced. Overall, this endeavor has provided a few more meronymic relations, reaching a number of 25.

3.1.4 Other Types of Relations: BabelNet was also used to provide other types of relations that are not present in PWN since they are not semantic relations such as hypernyms/hyponyms, and meronyms/holonyms. For instance, house and building, both concepts of the ontology, have in BabelNet a “purpose/use relation” with value shelter. The relation is not a semantic relation but a kind of relation that could be used in the ontology to establish interrelations among concepts in a meaningful way. Another example includes the following: earthquake is “cause of” disaster, landslide, and tsunami, all concepts of the ontology. Furthermore, earthquake is “a facet of” Earth. These relations from BabelNet further enrich the ontology schema.

3.1.5 Instances and Instantiation Issues: The ontology includes three instances, Earth, Northern Hemisphere and Western Hemisphere, with Earth being an instance of terrestrial planet, while the other two being instances of hemisphere. PWN is known to have considerable instantiation issues by including as concepts, lemmas that should be considered instances instead (Alfonseca and Manandhar, 2002, Miller and Hristea, 2006). For example, western and eastern hemispheres are not considered themselves instances but concepts in PWN. Indeed, in Wikipedia and BabelNet both of them are identified as instances. Hence, we decided to include them as such in the ontology, changing thus the original hierarchy.

3.2 Bottom-up ontology Enrichment and Population

The developed ontology has been used as a basis for the semantic information extraction process for a case study involving a corpus of 170 natural language texts (educational resources, news agencies articles, travel blogs, scientific articles) providing a wealth of spatial knowledge in terms of abstract spatial concepts and entities (place names).

Manual annotation of a subset of these texts revealed that they include various geospatial concepts and their instances that remained unexplored since they were not included in the ontology. Therefore a necessary second task of the approach aims at extracting additional semantic information from texts for the identification of either additional concepts (ontology enrichment) or instances of existing concepts (ontology population).

The task is supported by a tool implemented in R which guides domain experts in the identification of additional concepts and instances. The process is based on a semantic analysis of the text corpus for the identification of meaningful key phrases that may constitute a candidate for an additional concept or an instance of a concept. The spaCy Natural Language Processing software is used for the linguistic processing of the input corpus which includes: (a) tokenization to split text into words, phrases, symbols, or other meaningful elements called tokens, (b) sentence splitting to divide the texts into sentences, and (c) part-of-speech (POS) tagging to mark up each phrase as corresponding to a particular part of speech, and (d) lemmatization to identify the base or dictionary form of a word (lemma). Key phrases for the specific task are noun phrases consisting of a combination of adjectives, nouns (both common and proper), prepositions, and determiners as suggested by Handler et al., 2016. A further constraint is that noun phrases should contain one or more terms that refer to a concept from the ontology and are found in the body of the text at least twice. These noun phrases are identified using the phrasemachine software in R language. Figure 2 shows the result of

information extraction process. This includes locations (Dublin, Ireland, New England, etc.) using Named Entity Recognition, and ontology concepts (environment, place, geography, space, etc.) using extraction rules in the form of regular expressions and lexico-syntactic patterns. The list on the right part of Figure 2 shows noun phrases that could possibly refer to additional concepts or instances for enriching or populating the ontology (e.g., urban area, geography education, natural environment). Among the identified noun phrases, the most prominent ones for constituting candidate ontology concepts are:

- Noun phrases that form PWN lemmas such as urban area, magnetic north, and compass point
- Noun phrases that form BabelNet lemmas, such as mental map, nautical chart, and urban environment
- Noun phrases that are comprised by existing ontology concepts, such as map scale, community centre, time zone.

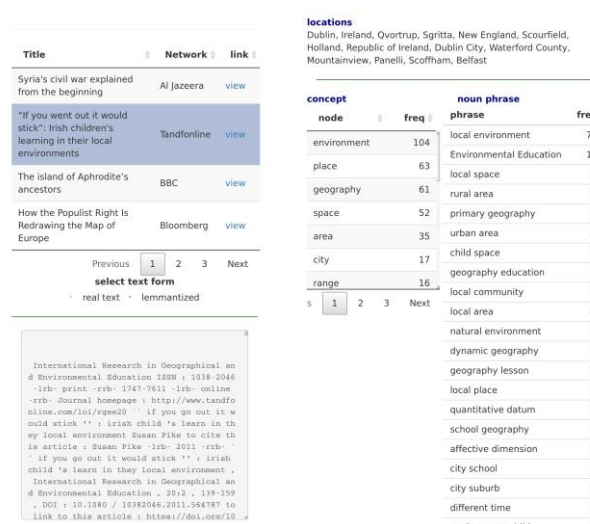


Figure 2. Extraction of locations, spatial concepts, and noun phrases for identifying additional concepts and instances

For example, the linguistic analysis of the test corpus identified the following city-related concepts: city centre, capital city, metropolitan city, and ancient city, as well as instances such as city of Athens, city of Berlin, Paris, etc.

Noun phrases however, in a more semantic approach can be used to extract semantic relations among concepts. Work by Nastase et al. (2006), has indicated 30 classes of semantic relations organised in five superclasses; causality, temporality, spatial, participant and quality. Thus, analysis of noun phrases could indicate relations among concepts such as in the case of nautical chart, which could be a hyponym of the concepts chart and map, but on the other hand, denotes a purpose/use relation (causality); a chart used for navigation. A similar case is desert storm, where the relation located at/in holds between the two concepts (a storm taking place in the desert).

Noun phrases can also be used for ontology population and this highlights again the issue discussed in 1.1.5. Different methods have been proposed to identify instances among the Wikipedia content (Zirn et al., 2008) and within PWN (Alfonseca and Manandhar, 2002, Miller and Hristea, 2006). These methods can also be applied in our case to differentiate between concepts and instances, using Named Entity Recognition to identify

locations, organizations, and persons (e.g. British Museum, White House).

Moreover, lexico-syntactic patterns could also be of use. For example, in Wikipedia the instantiation of earthquakes follows a specific pattern: “(Date) Year Location earthquake”, e.g. April 2011 Fukushima earthquake. Including similar patterns to the process of information extraction will provide instances to the concepts of the ontology. Indeed, the process allowed us to extract the 1953 Cephalonia earthquake as instance of earthquake mentioned in the test text corpus.

Figure 3 shows an example of the enrichment and population process for the concept “earthquake”. Circles represent concepts and rectangles instances. Blue shapes correspond to the initial ontology concepts and instances. Orange rectangles refer to BabelNet derived instances, whereas the purple rectangle is an instance extracted through the information extraction process. Pink circles on the left are synonyms of the concept earthquake that could be used as alternative mentions of the ontology concept during semantic search.

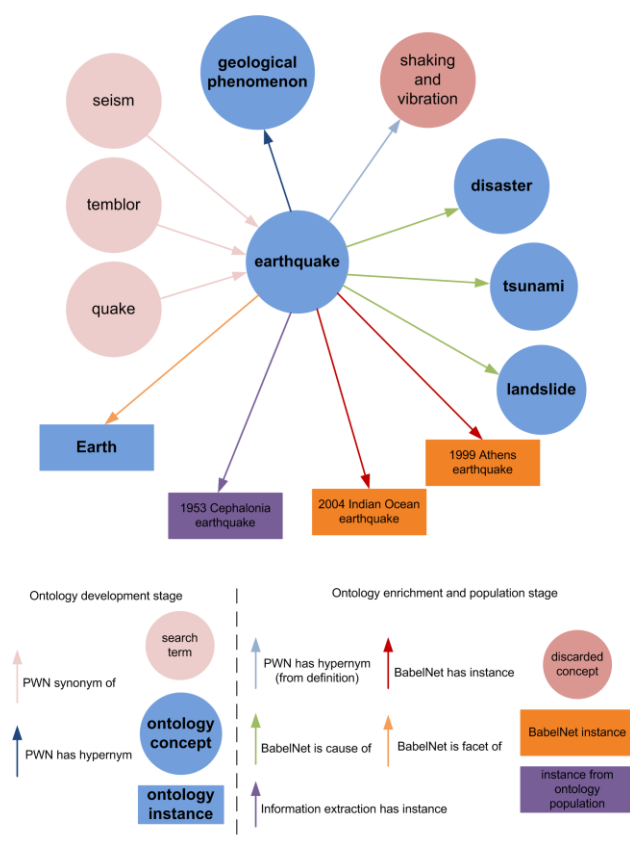


Figure 3. An example of ontology enrichment and population

4. DISCUSSION AND CONCLUSIONS

The paper describes an approach for enriching and populating a geospatial ontology based on existing ontologies and natural language texts. The ontology may guide semantic information extraction from textual content and support semantic search and exploration of texts.

A significant challenge faced by the approach and generally by ontology-based information extraction approaches is word-sense disambiguation. The latent polysemy and ambiguity in natural language texts cannot be resolved solely based on a generic or domain ontology with a natural language anchorage. As already mentioned in Section 3, PWN synsets include terms which are not unique and induce difficulties during their identification. Word-sense disambiguation is a critical research field for improving the quality of information extraction process and will be considered as a future step of the approach.

Future work will also consider formulating additional patterns for extracting instances from natural language semi-structured or unstructured texts. A critical issue in ontology development is the differentiation between concepts and instances in a consistent and semantically solid and meaningful way. Since we are not interested only in place names extracted by NER, which nowadays perform quite well, but our interest covers different instantiations related to the geospace, instance extraction and subsequent ontology population is an endeavor worth pursuing.

Additionally, the overall process as presented herein will be tested in larger corpora from different domains to confirm (a) validity and (b) portability of the approach. Corpora from other domains can also provide insights of the spatiality included therein and how geospace interacts with other domains to extract latent relations between them.

REFERENCES

- Alfonseca, E., and Manandhar, S., 2002. Distinguishing concepts and instances in WordNet. In: *Proceedings of the First International Conference of Global WordNet Association*, Mysore, India.
- Ballatore, A., and Adams, B., 2015. Extracting Place Emotions from Travel Blogs. In: *Proceedings of AGILE 2015 Geographic Information Science as an Enabler of Smarter Cities and Communities*, Lecture Notes in Geoinformation and Cartography, F. Bacao, M. Y. Santos, and M. Painho, Eds. Springer, pp 1–5.
- Bruggmann, A., and Fabrikant, S.I., 2014a. Spatializing time in a history text corpus. In: *GIScience 2014: Eighth International Conference on Geographic Information Science*, Vienna (A), 23–26 September 2014, pp 183–186.
- Bruggmann, A., and Fabrikant, S.I., 2014b. How to visualize the geography of Swiss history. In *Proceedings of the AGILE 2014 International Conference on Geographic Information Science: Connecting a Digital Europe Through Location and Place*, J. J. Huerta, S. Schade, C. Granel. (Eds.), Castellón, June, 3–6, 2014, https://agile-online.org/conference_paper/cds/agile_2014/agile2014_144.pdf (13 June 2018).
- Cooper, D.C., Gregory, I.N., Hardie, A., and Rayson, P., 2015. Spatializing and Analyzing Digital Texts: Corpora, GIS, and Places. In: *Deep Maps and Spatial Narratives*. Indiana University Press/Bloomington, Indiana.
- Derungs, C., and Purves, R.S., 2014. From text to landscape: locating, identifying and mapping the use of landscape features

- in a Swiss Alpine corpus. *International Journal of Geographical Information Science* 28, 6, pp 1272-1293.
- Fellbaum C., 1998 (ed.) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Fernandez, M., Gomez-Perez, A. and Juristo, N., 1997 "METHONTOLOGY: From Ontological Art towards Ontological Engineering", AAAI-97 Spring Symposium on Ontological Engineering, Stanford University, March 24-26.
- Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A., 2003. Sweetening WordNet with DOLCE. *AI Magazine* 24(3), pp. 13–24.
- GeoNames. GeoNames. <http://geonames.org/> (13 July 2018).
- Guarino, N., and Welty, C., 2004. An overview of OntoClean. In *Handbook on Ontologies*, S. Staab and R. Studer (Eds.) Springer-Verlag, pp. 151–159.
- Handler, A., Denny, M. J., Wallach, H., O'Connor, B., 2016. Bag of What? Simple Noun Phrase Extraction for Text Analysis. In *Proceedings of the Workshop on Natural Language Processing and Computational Social Science* at the 2016 Conference on Empirical Methods in Natural Language Processing.
- Hu, Y., Janowicz, K., Prasad, S., Gao, S., 2015. Enabling Semantic Search and Knowledge Discovery for ArcGIS Online: A Linked-Data-Driven Approach. In: *Proceedings of AGILE 2015 Geographic Information Science as an Enabler of Smarter Cities and Communities*, Lecture Notes in Geoinformation and Cartography, F. Bacao, M. Y. Santos, and M. Painho, Eds. Springer, pp 107-124.
- Jones, C.B., Alani, H., Tudhope, D., 2001. Geographical Information Retrieval with Ontologies of Place. In: *Spatial Information Theory. COSIT 2001*, Lecture Notes in Computer Science, Montello D.R. (ed) vol 2205. Springer, Berlin, Heidelberg.
- Jones, C.B., and Purves R.S., 2008. Geographical information retrieval. *International Journal of Geographical Information Science*, 22:3, pp. 219-228,
- Kaplan, A.N., and Schubert, L.K., 2001. Measuring and improving the quality of world knowledge extracted from WordNet, Tech. Rep. 751, Dept. of Computer Science, Univ. of Rochester, Rochester, NY 14627-0226.
- Karkaletsis, V., Fragkou, P., Petasis, G., and Iosif, E., 2011. Ontology Based Information Extraction from Text. In *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, G. Paliouras, C.D. Spyropoulos, and G. Tsatsaronis (Eds.) Lecture Notes in Computer Science, vol 6050. Springer, Berlin, Heidelberg, pp. 89-109.
- Kavouras, M., Kokla, M., Tomai, E., Darra, A., Pastra, K., 2016. GEOTHNK: A Semantic Approach to Spatial Thinking. In: *Progress in Cartography: EuroCarto 2015*, Gartner, G., Jobst, M., Huang, H., (Eds.), Springer, pp. 319–338.
- Kokla M., Tomai E., and Papadias V., 2018. Educating spatial knowledge from text: semantic information extraction of concepts and places. *Journal of Spatial Information Science* (under review).
- Lutz, M. and Klien, E., 2006. Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20:3, pp 233-260.
- McCrae, J. P., and Prangnawarat, N., 2016. Identifying Poorly-Defined Concepts in WordNet with Graph Metrics. In: *Knowledge Graphs and Language Technology, ISWC 2016 International Workshops: KEKI and NLP & DBpedia*, van Erp, M., Hellmann, S., McCrae, J.P., Chiacos, C., Choi, K.-S., Gracia, J., Hayashi, Y., Koide, S., Mendes, P., Paulheim, H., Takeda, H. (Eds.), pp. 66-75.
- Miller, G.A., and Hristea F., 2006. WordNet Nouns: Classes and Instances. *Comput. Linguist.* 32, 1, pp. 1-3. DOI=<http://dx.doi.org/10.1162/coli.2006.32.1.1>
- Nastase, V., Sayyad-Shirabad, J., Sokolova, M., and Szpakowicz S., 2006. Learning noun-modifier semantic relations with corpus-based and Wordnet-based features. In: *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, Boston, Massachusetts, pp. 781-787.
- Navigli, R. and Ponzetto, S., 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence*, 193, Elsevier, pp. 217-250.
- Nédellec, C., and Nazarenko, A., 2005. Ontologies and Information Extraction. LIPN Internal Report, <https://hal-univ-paris13.archives-ouvertes.fr/hal-00098068/document> (13 July 2018).
- Petasis, G., Karkaletsis, V., Paliouras, G., Krithara A., Zavitsanos. E., 2011. Ontology Population and Enrichment: State of the Art In: *Multimedia Information Extraction*, G. Paliouras et al. (Eds.), LNAI 6050, Springer-Verlag Berlin Heidelberg, pp. 134–166.
- Wimalasuriya, D.C., and Dou, D., 2010. Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science* 36, 3, pp. 306 – 323.
- Zirn, C., Nastase, V., and Strube M., 2008. Distinguishing between instances and classes in the wikipedia taxonomy. In: *Proceedings of the 5th European semantic web conference on The semantic web: research and applications (ESWC'08)*, S. Bechhofer, M. Hauswirth, J. Hoffmann, and M.Koubarakis (Eds.). Springer-Verlag, Berlin, Heidelberg, pp. 376-387.